Full wwPDB NMR Structure Validation Report

Mar 8, 2026 – 09:27 AM UTC

PDB ID	:	2KP6 / pdb_00002kp6
BMRB ID	:	16545
Title	:	Solution NMR structure of protein CV0237 from Chromobacterium violaceum. Northeast Structural Genomics Consortium (NESG) target CvT1
Authors	:	Fares, C.; Lemak, A.; Yee, A.; Garcia, M.; Montelione, G.T.; Arrowsmith, C.T.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2009-10-07

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

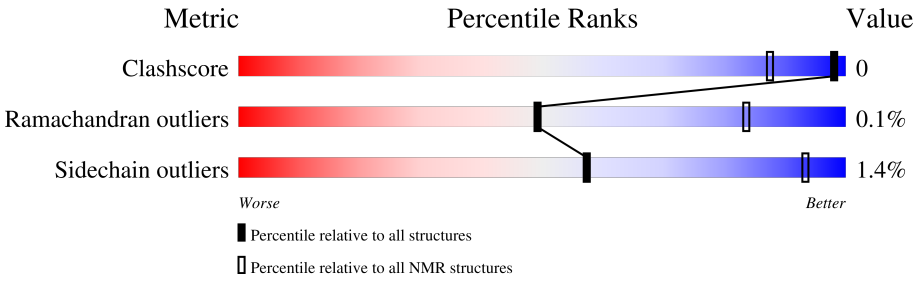
MolProbity	:	4-5-2 with Phenix2.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 85%.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	229148	14424
Ramachandran outliers	224038	12848
Sidechain outliers	223484	12823

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and a grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	82	85% • 10% •

2 Ensemble composition and analysis

This entry contains 20 models. Model 12 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:6-A:76 (71)	0.68	12

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 4 clusters and 5 single-model clusters were found.

Cluster number	Models
1	2, 6, 11, 12, 17, 20
2	1, 3, 7, 8, 13
3	4, 19
4	5, 18
Single-model clusters	9; 10; 14; 15; 16

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1196 atoms, of which 587 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Uncharacterized protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	79	1196	389	587	103	115	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLN	-	expression tag	UNP Q7P1H6
A	-1	GLY	-	expression tag	UNP Q7P1H6
A	0	HIS	-	expression tag	UNP Q7P1H6
A	69	GLY	GLU	engineered mutation	UNP Q7P1H6

4 Residue-property plots

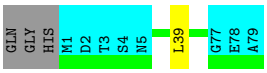
4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: Uncharacterized protein

Chain A:





4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Uncharacterized protein

Chain A:



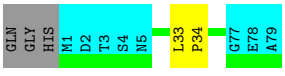


4.2.2 Score per residue for model 2

• Molecule 1: Uncharacterized protein

Chain A:

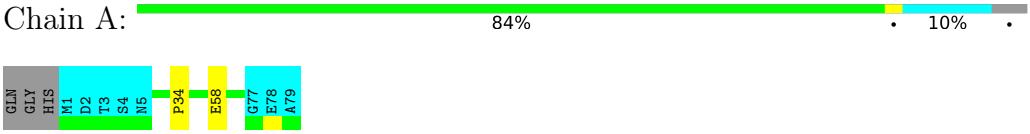




4.2.3 Score per residue for model 3

• Molecule 1: Uncharacterized protein

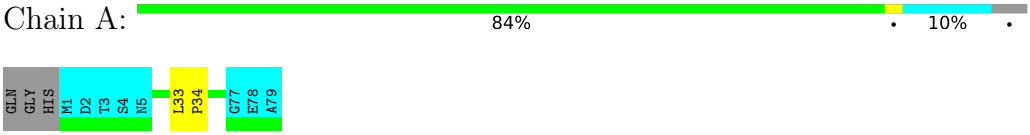
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Uncharacterized protein

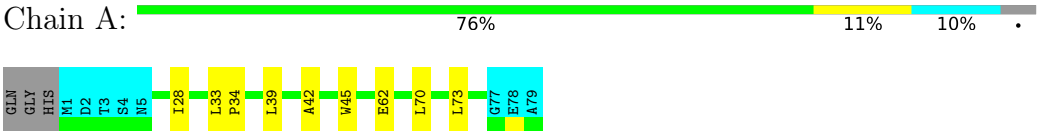
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Uncharacterized protein

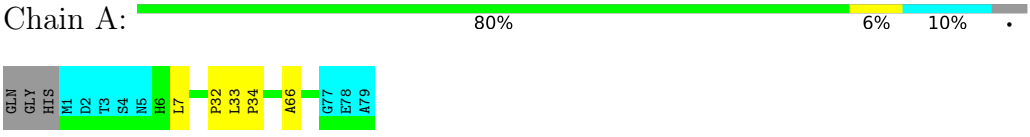
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Uncharacterized protein

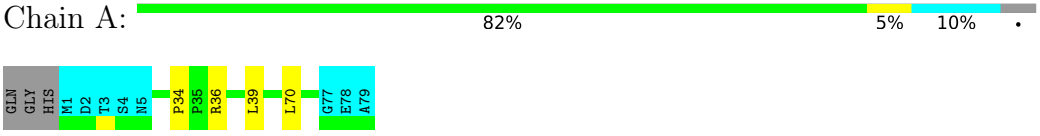
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Uncharacterized protein

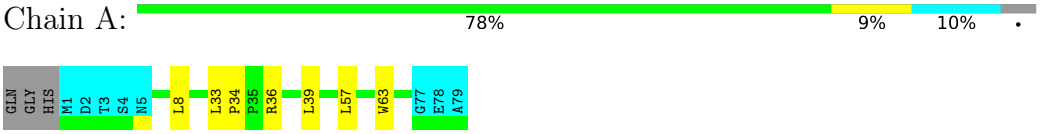
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Uncharacterized protein

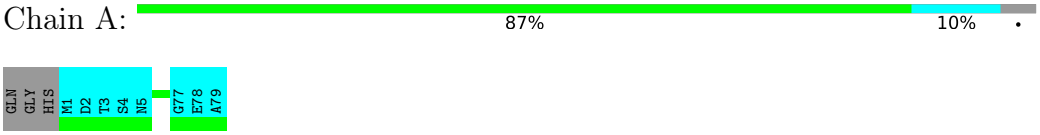
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Uncharacterized protein

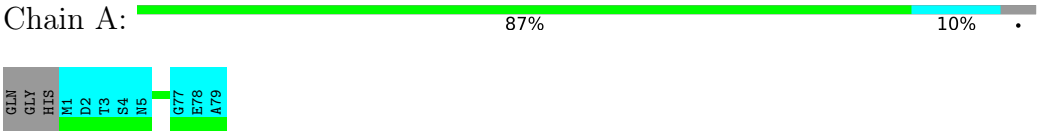
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Uncharacterized protein

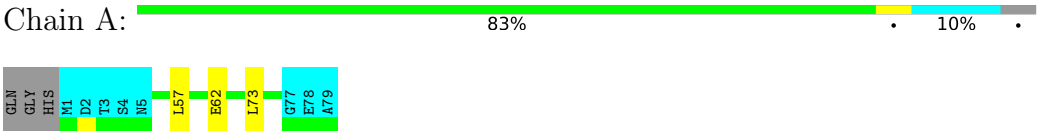
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Uncharacterized protein

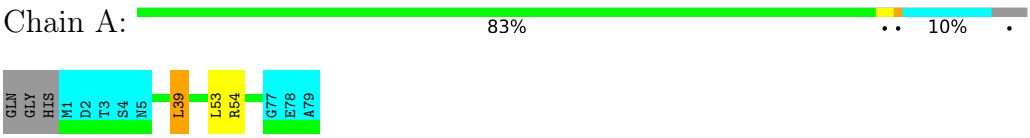
Chain A:



4.2.12 Score per residue for model 12 (medoid)

• Molecule 1: Uncharacterized protein

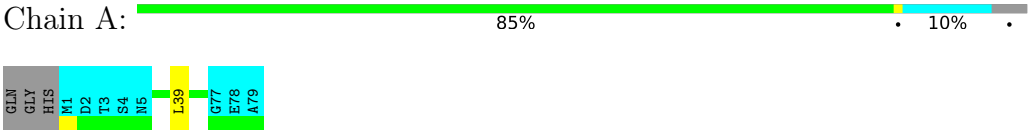
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Uncharacterized protein

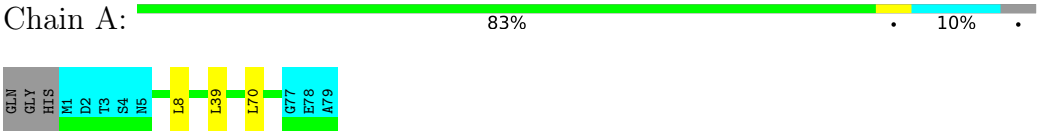
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Uncharacterized protein

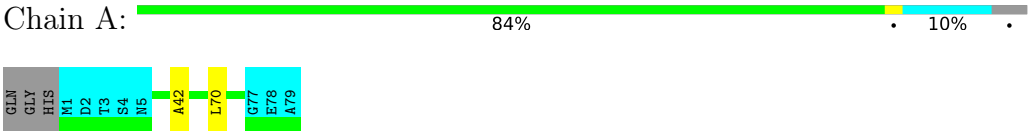
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Uncharacterized protein

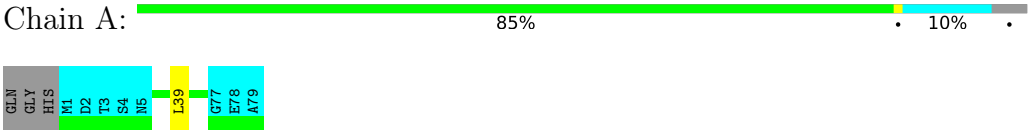
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Uncharacterized protein

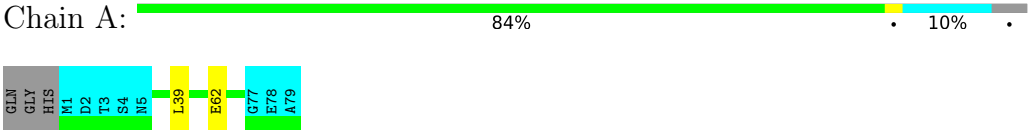
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Uncharacterized protein

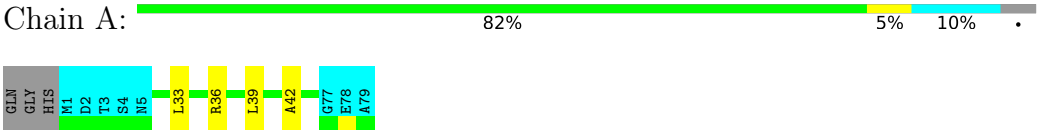
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Uncharacterized protein

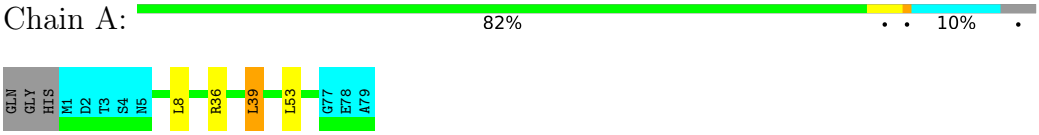
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Uncharacterized protein

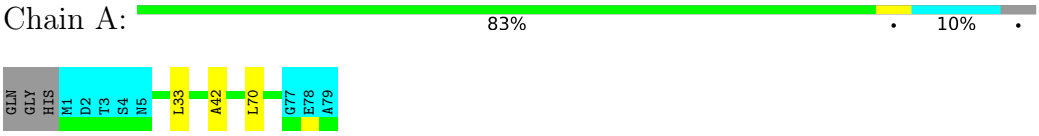
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Uncharacterized protein

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *molecular dynamics*.

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *all calculated structures submitted*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
CNS	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	869
Number of shifts mapped to atoms	869
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

6 Model quality

6.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	1.28±0.04	1±1/572 (0.2± 0.2%)	1.03±0.03	0±0/787 (0.0± 0.0%)
All	All	1.28	25/11440 (0.2%)	1.03	0/15740 (0.0%)

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	34	PRO	CA-C	6.29	1.55	1.51	4	7
1	A	39	LEU	C-N	6.11	1.39	1.33	7	11
1	A	33	LEU	C-N	5.39	1.38	1.33	4	5
1	A	42	ALA	C-N	5.12	1.38	1.33	15	1
1	A	8	LEU	C-N	5.08	1.41	1.33	19	1

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	553	542	540	0±1
All	All	11060	10840	10800	9

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:39:LEU:HG	1:A:53:LEU:HD13	0.49	1.85	12	2
1:A:33:LEU:HD11	1:A:42:ALA:HB2	0.48	1.83	20	2
1:A:7:LEU:HB3	1:A:66:ALA:HA	0.47	1.85	6	1
1:A:8:LEU:HD11	1:A:70:LEU:HD13	0.45	1.89	14	1
1:A:34:PRO:HB2	1:A:36:ARG:HG2	0.43	1.91	7	1
1:A:28:ILE:HD12	1:A:73:LEU:HB3	0.42	1.91	5	1
1:A:42:ALA:HB3	1:A:45:TRP:CD1	0.40	2.51	5	1

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	71/82 (87%)	69±1 (98±2%)	2±1 (2±2%)	0±0 (0±0%)	49	83
All	All	1420/1640 (87%)	1387 (98%)	32 (2%)	1 (0%)	49	83

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	32	PRO	1

6.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	56/64 (88%)	55±1 (99±1%)	1±1 (1±1%)	57	93
All	All	1120/1280 (88%)	1104 (99%)	16 (1%)	57	93

All 7 unique residues with a non-rotameric sidechain are listed below. They are sorted by the

frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	62	GLU	4
1	A	70	LEU	4
1	A	36	ARG	3
1	A	57	LEU	2
1	A	58	GLU	1
1	A	63	TRP	1
1	A	54	ARG	1

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates

There are no oligosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 85% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	869
Number of shifts mapped to atoms	869
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	77	-0.27 ± 0.20	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	73	0.35 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	62	-0.33 ± 0.19	None needed (< 0.5 ppm)
^{15}N	62	0.54 ± 0.36	None needed (imprecise)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 85%, i.e. 823 atoms were assigned a chemical shift out of a possible 970. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	323/340 (95%)	133/136 (98%)	131/142 (92%)	59/62 (95%)
Sidechain	480/550 (87%)	329/362 (91%)	151/171 (88%)	0/17 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	20/80 (25%)	10/40 (25%)	8/34 (24%)	2/6 (33%)
Overall	823/970 (85%)	472/538 (88%)	290/347 (84%)	61/85 (72%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 83%, i.e. 869 atoms were assigned a chemical shift out of a possible 1052. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	344/381 (90%)	143/153 (93%)	139/158 (88%)	62/70 (89%)
Sidechain	505/591 (85%)	346/388 (89%)	159/185 (86%)	0/18 (0%)
Aromatic	20/80 (25%)	10/40 (25%)	8/34 (24%)	2/6 (33%)
Overall	869/1052 (83%)	499/581 (86%)	306/377 (81%)	64/94 (68%)

7.1.4 Statistically unusual chemical shifts (i)

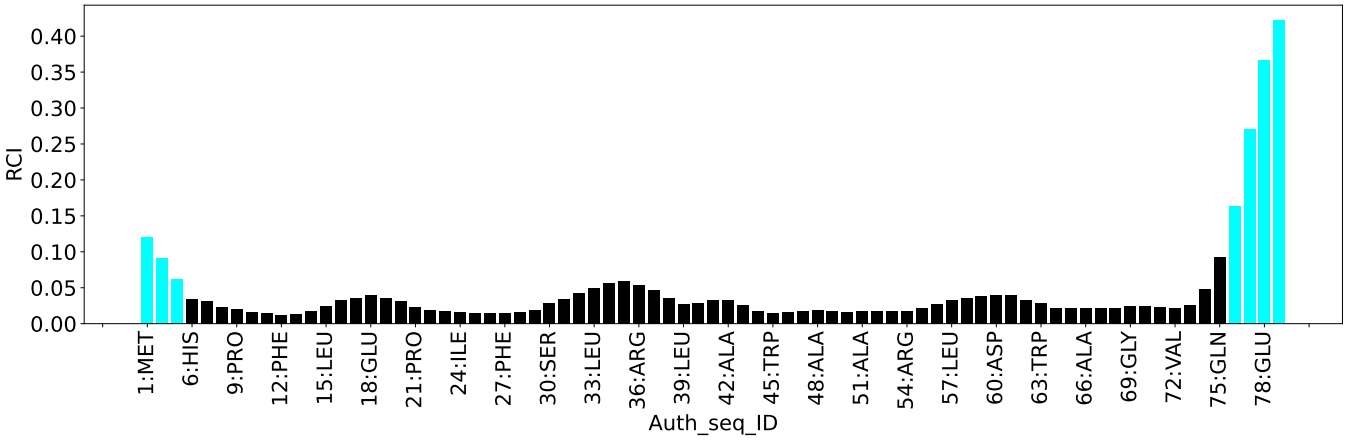
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	14	GLN	HG3	0.90	0.91 – 3.68	-5.0

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:


RCI
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1:MET
6:HIS
9:PRO
12:PHE
15:LEU
18:GLU
21:PRO
24:ILE
27:PHE
30:SER
33:LEU
36:ARG
39:LEU
42:ALA
45:TRP
48:ALA
51:ALA
54:ARG
57:LEU
60:ASP
63:TRP
66:ALA
69:GLY
72:VAL
75:GLN
78:GLU
Auth_seq_ID